

# UTILITY PATENT APPLICATION TRANSMITTAL (Large Entity)

(Only for new nonprovisional applications under 37 CFR 1.53(b))

Docket No.  
13220

Total Pages in this Submission

## TO THE ASSISTANT COMMISSIONER FOR PATENTS

Box Patent Application  
Washington, D.C. 20231

Transmitted herewith for filing under 35 U.S.C. 111(a) and 37 C.F.R. 1.53(b) is a new utility patent application for an invention entitled:

**CONTINUOUS SPEECH RECOGNITION APPARATUS AND METHOD**

and invented by:

**Takashi Tomoeda**

1c503 U.S. PTO  
09/453153  
12/02/99

If a CONTINUATION APPLICATION, check appropriate box and supply the requisite information:

☐ Continuation ☐ Divisional ☐ Continuation-in-part (CIP) of prior application No.: \_\_\_\_\_

Which is a:

☐ Continuation ☐ Divisional ☐ Continuation-in-part (CIP) of prior application No.: \_\_\_\_\_

Which is a:

☐ Continuation ☐ Divisional ☐ Continuation-in-part (CIP) of prior application No.: \_\_\_\_\_

Enclosed are:

### Application Elements

1. ☒ Filing fee as calculated and transmitted as described below
2. ☒ Specification having 28 pages and including the following:
  - a. ☒ Descriptive Title of the Invention
  - b. ☐ Cross References to Related Applications (if applicable)
  - c. ☐ Statement Regarding Federally-sponsored Research/Development (if applicable)
  - d. ☐ Reference to Microfiche Appendix (if applicable)
  - e. ☒ Background of the Invention
  - f. ☒ Brief Summary of the Invention
  - g. ☒ Brief Description of the Drawings (if drawings filed)
  - h. ☒ Detailed Description
  - i. ☒ Claim(s) as Classified Below
  - j. ☒ Abstract of the Disclosure

**UTILITY PATENT APPLICATION TRANSMITTAL**  
**(Large Entity)**

*(Only for new nonprovisional applications under 37 CFR 1.53(b))*

Docket No.

13220

Total Pages in this Submission

**Application Elements (Continued)**

3. ☒ Drawing(s) *(when necessary as prescribed by 35 USC 113)*
- a. ☒ Formal                      Number of Sheets 4
- b. ☐ Informal                      Number of Sheets \_\_\_\_\_
4. ☒ Oath or Declaration
- a. ☒ Newly executed *(original or copy)*                      ☐ Unexecuted
- b. ☐ Copy from a prior application (37 CFR 1.63(d)) *(for continuation/divisional application only)*
- c. ☒ With Power of Attorney                      ☐ Without Power of Attorney
- d. ☐ DELETION OF INVENTOR(S)  
Signed statement attached deleting inventor(s) named in the prior application,  
see 37 C.F.R. 1.63(d)(2) and 1.33(b).
5. ☐ Incorporation By Reference *(usable if Box 4b is checked)*  
The entire disclosure of the prior application, from which a copy of the oath or declaration is supplied  
under Box 4b, is considered as being part of the disclosure of the accompanying application and is hereby  
incorporated by reference therein.
6. ☐ Computer Program in Microfiche *(Appendix)*
7. ☐ Nucleotide and/or Amino Acid Sequence Submission *(if applicable, all must be included)*
- a. ☐ Paper Copy
- b. ☐ Computer Readable Copy *(identical to computer copy)*
- c. ☐ Statement Verifying Identical Paper and Computer Readable Copy

**Accompanying Application Parts**

8. ☒ Assignment Papers *(cover sheet & document(s))*
9. ☐ 37 CFR 3.73(B) Statement *(when there is an assignee)*
10. ☐ English Translation Document *(if applicable)*
11. ☒ Information Disclosure Statement/PTO-1449                      ☒ Copies of IDS Citations
12. ☐ Preliminary Amendment
13. ☒ Acknowledgment postcard
14. ☒ Certificate of Mailing
- ☐ First Class                      ☒ Express Mail *(Specify Label No.):* EM341609001US

**UTILITY PATENT APPLICATION TRANSMITTAL**  
**(Large Entity)**

*(Only for new nonprovisional applications under 37 CFR 1.53(b))*

Docket No.  
13220

Total Pages in this Submission

**Accompanying Application Parts (Continued)**

15. ☒ Certified Copy of Priority Document(s) *(if foreign priority is claimed)*

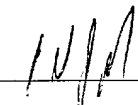
16. ☐ Additional Enclosures *(please identify below):*

**Fee Calculation and Transmittal**

**CLAIMS AS FILED**

For	#Filed	#Allowed	#Extra	Rate	Fee
Total Claims	10	- 20 =	0	x \$18.00	\$0.00
Indep. Claims	5	- 3 =	2	x \$78.00	\$156.00
Multiple Dependent Claims (check if applicable) <input type="checkbox"/>					\$0.00
BASIC FEE					\$760.00
OTHER FEE (specify purpose) <u>Assignment recordal fee</u>					\$40.00
TOTAL FILING FEE					\$956.00

- ☒ A check in the amount of \$956.00 to cover the filing fee is enclosed.
- ☒ The Commissioner is hereby authorized to charge and credit Deposit Account No. 19-1013/SSMP as described below. A duplicate copy of this sheet is enclosed.
- ☐ Charge the amount of \_\_\_\_\_ as filing fee.
- ☒ Credit any overpayment.
- ☒ Charge any additional filing fees required under 37 C.F.R. 1.16 and 1.17.
- ☐ Charge the issue fee set in 37 C.F.R. 1.18 at the mailing of the Notice of Allowance, pursuant to 37 C.F.R. 1.311(b).



Signature

Paul J. Esatto, Jr.  
Registration No. 30,749  
SCULLY, SCOTT, MURPHY & PRESSER  
400 Garden City Plaza  
Garden City, NY 11530  
(516) 742-4343

Dated: December 2, 1999

CC:

- 1 -

## CONTINUOUS SPEECH RECOGNITION APPARATUS AND METHOD

BACKGROUND OF THE INVENTIONField of the Invention

5           The present invention relates to a continuous speech recognition apparatus and method, and more particularly to a continuous speech recognition apparatus and method which achieves augmentation in speed and accuracy of recognition.

Description of the Related Art

10           As an example of a conventional continuous speech recognition apparatus, reference is had to a paper by S. Ortmanns, "LANGUAGE-MODEL LOOK-AHEAD FOR LARGE VOCABULARY SPEECH RECOGNITION", ICSLP, 1996.

15           The conventional continuous speech recognition apparatus is shown in FIG. 6. Referring to FIG. 6, the conventional continuous speech recognition apparatus shown includes a hypothesis storage section 1, a hypothesis expansion section 3, a tree structure dictionary storage section 4, a language model section 7, and an acoustic model section 8.

20           In operation, the hypothesis storage section 1 stores hypotheses therein. The tree structure dictionary storage section 4 stores words, which make an object of recognition, as a tree structure dictionary (refer to FIG. 2). The acoustic model section 8 calculates an acoustic model score for each  
25           frame. The language model section 7 calculates a language model score.

The hypothesis expansion section 3 acquires, for each frame, a structure of arcs from the tree structure dictionary storage section 4 taking an acoustic model from the acoustic model section 8 and a language model score from the language model section 7 into consideration and expands a hypothesis present on an arc to a succeeding arc. Referring to FIG. 2, a tree structure dictionary is structured such that a word is reached by tracing arcs branching in a tree structure from a root to a leaf (terminal arc).

Speech which makes an object of recognition is divided into short-time frames of a predetermined period, and such expansion as described above (that is, expansion of a hypothesis on an arc of a tree structure dictionary to a succeeding arc) is repeated from the speech beginning frame to the speech terminating frame. Then, a word through which a hypothesis which exhibits the highest score has passed in the past (a terminal of the tree structure dictionary) is finally determined as a recognition result.

Here, a hypothesis has position information of an arc on a tree structure dictionary, a history until the position is reached, and a score.

In a continuous speech recognition system wherein a plurality of words are represented as one tree structure dictionary (refer to FIG. 2), what is a word with regard to which a hypothesis is being expanded at present cannot be specified except at the terminal arc.

Therefore, although an acoustic model score is calculated for each frame, a language model score can originally be determined only when a hypothesis reaches a terminal arc of a tree structure dictionary.

5           Therefore, in order to add a language model score as early as possible, a method employing look-ahead of a unigram language model score and look-ahead of a bigram language model score is disclosed in the document mentioned hereinabove.

10           According to the look-ahead of a unigram language model score, the highest one of unigram language model scores of words settled at terminal arcs in a tree structure dictionary is provided to a predecessor arc, and the unigram language model provided to the arc is temporarily added as a language model score of the hypothesis present on the arc, and then, when the  
15           hypothesis reaches the terminal arc of the tree structure dictionary and the word is settled, the unigram language model score which has been used till then is abandoned and then the settled bigram language model score is added.

20           On the other hand, according to the look-ahead of a bigram language model score, when a context is determined and a new tree structure dictionary is produced, bigram language model scores regarding all words of the context are calculated, and that one of the language model scores which has the highest score is provided to a predecessor arc, and then the bigram  
25           language score provided to the arc is added as a language model score of the hypothesis present on a certain arc.

The conventional speech recognition system has the following problems.

The first problem resides in that, when look-ahead of a bigram language model score is performed, a great memory capacity and a large amount of calculation are required.

The reason is that, where look-ahead of a bigram language model score is performed, when a context is produced and a tree structure dictionary is produced newly, it is required to repeat processing of producing not part of a tree structure dictionary but an entire tree structure dictionary, calculating all bigram language model scores with respect to the context and provide language model scores of all terminal arcs in the tree structure dictionary, with which words are settled, to a predecessor arc to propagate the language model scores to all predecessor arcs.

The second problem resides in that, when look-ahead of a unigram language model score is performed, wasteful calculation is performed.

The reason is that, when look-ahead of a unigram language model score is performed, some of arcs of a tree structure dictionary may expand only to a word whose connection to the context is not permitted linguistically and the hypothesis is expanded also to such arc, in which wasteful calculation is involved.

The third problem is such as follows. If strict look-ahead of a language model score of a bigram or more is

not performed using a frame synchronous beam search (for the frame synchronous beam search, for example, Hermann Ney, "Data Driven Search Organization for Continuous Speech Recognition", IEEE TRANSACTIONS ON SIGNAL PROCESSING, February, 1992 is referred to), that is, if connection possibility according to linguistic restrictions between a context and a word in a tree structure dictionary is not looked ahead, then the hypothesis is expanded also to an arc which is developed to a word whose connection to a context is not permitted linguistically as described above in connection with the second problem.

Then, if the score of the hypothesis is much higher than the others, then all hypotheses on an arc which is developed to a word whose connection to the context is permitted linguistically are excluded from the beam and thus eliminated.

As a result, in the succeeding frames, the word cannot be connected to a next word at all, and recognition processing for speech uttered later is disabled. In other words, recognition processing cannot be performed any more and a recognition result cannot be outputted.

#### SUMMARY OF THE INVENTION

It is an object of the present invention to provide a continuous speech recognition apparatus and method by which the recognition speed and the recognition accuracy in continuous speech recognition can be augmented.

In order to attain the object described above, according



to an aspect of the present invention, there is provided a continuous speech recognition apparatus, comprising a hypothesis storage section for storing hypotheses therein, hypothesis expansion discrimination means for determining  
5 whether or not a hypothesis may be expanded to a succeeding arc, a tree structure dictionary storage section for storing a tree structure dictionary and a context preceding to the tree structure dictionary therein, a succeeding word speech part information storage section for storing information of whether  
10 or not speech parts are included in all of succeeding words present behind each of arcs in the tree structure dictionary, a speech part connection information storage section for storing connection information between the speech parts, means for providing a language model score to a hypothesis, means  
15 for providing an acoustic model score to a hypothesis, and hypothesis expansion means operable in response to an expansion instruction received from the hypothesis expansion discrimination means for acquiring a structure of an arc from the tree structure dictionary storage section and expanding  
20 a hypothesis present on the arc to a succeeding arc taking the acoustic model score and the language model score into consideration and then storing a result of the expansion into the hypothesis storage section.

According to another aspect of the present invention,  
25 there is provided a continuous speech recognition method for a continuous speech recognition apparatus which includes a

hypothesis storage section for storing hypotheses therein, a  
tree structure dictionary storage section for storing a tree  
structure dictionary and a context preceding to the tree  
structure dictionary therein, a succeeding word speech part  
5 information storage section for storing information of whether  
or not speech parts are included in all of succeeding words  
present behind each of arcs in the tree structure dictionary,  
and a speech part connection information storage section for  
storing connection information between the speech parts,  
10 comprising the step of repeating a process for all of hypotheses  
present at a certain frame time, the process including the steps  
of acquiring a context of a tree structure dictionary to which  
a hypothesis belongs from the tree structure dictionary storage  
section, acquiring speech part connection information of the  
15 speech parts of the context from the speech part connection  
information storage section, acquiring arcs in the tree  
structure dictionary to which the hypothesis belongs from the  
hypothesis storage section, and repeating, for all succeeding  
arcs immediately succeeding the arcs, a process including the  
20 steps of acquiring, where an arc selected at present is  
represented as first arc and a succeeding arc immediately  
succeeding the first arc is represented as second arc,  
succeeding word speech part information of the second arc from  
the succeeding word speech part information storage section,  
25 discriminating from the acquired speech part connection  
information and the acquired succeeding word speech part

information whether or not the hypothesis may be expanded from the first arc to the second arc and determining that the hypothesis must not be expanded to the second arc if a connectable speech part included in the speech part connection information is not detected behind the second arc, but determining otherwise that the hypothesis may be expanded to the second arc, expanding the hypothesis to the second arc. and discriminating whether or not the loop has been completed for all of the hypotheses and ending, when the loop has been completed for all of the hypotheses, the expanding processing of the hypotheses of the frame in a frame synchronous beam search.

With the continuous speech recognition apparatus and method, a hypothesis is prevented from being expanded to an arc to a word which cannot connect linguistically to a context. Consequently, the number of unnecessary hypotheses is minimized and the speed of continuous speech recognition is augmented as much. Further, occurrence of a situation that the score of a hypothesis to a word which cannot connect linguistically to a context is higher than those of the other hypotheses is prevented. Consequently, the recognition accuracy in continuous speech recognition is augmented.

The above and other objects, features and advantages of the present invention will become apparent from the following description and the appended claims, taken in conjunction with the accompanying drawings in which like parts or elements are

denoted by like reference symbols.

#### BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a block diagram of a continuous speech  
5 recognition apparatus to which the present invention is  
applied;

FIG. 2 is a diagrammatic view schematically showing an  
example of a tree structure dictionary used in the continuous  
speech recognition apparatus of FIG. 1;

10 FIG. 3 is a diagrammatic view illustrating an example  
of succeeding word speech part information in the continuous  
speech recognition apparatus of FIG. 1;

FIG. 4 is a diagrammatic view showing an example of speech  
part connection information to a particle used in the  
15 continuous speech recognition apparatus of FIG. 1;

FIG. 5 is a flow chart illustrating operation of the  
continuous speech recognition apparatus of FIG. 1; and

FIG. 6 is diagrammatic view showing a conventional speech  
recognition apparatus.

20

#### DESCRIPTION OF THE PREFERRED EMBODIMENT

FIG. 1 shows a continuous speech recognition apparatus  
to which the present invention is applied. The continuous  
speech recognition apparatus uses a frame synchronous beam  
25 search. In the following description, only operation of  
expansion of a hypothesis for one frame in a frame synchronous

beam search is described. The other operation of the continuous speech recognition apparatus conforms with that of a frame synchronous beam search.

Referring to FIG. 1, the continuous speech recognition apparatus shown includes a hypothesis storage section 1 for storing hypotheses to be used in a search of continuous speech recognition, and a tree structure dictionary storage section 4 for storing contexts and all words which make an object of recognition as tree structure dictionaries (refer to FIG. 2). The tree structure dictionaries are stored separately for the individual contexts.

A succeeding word speech part information storage section 5 stores speech part information of all succeeding words of each arc of a tree structure dictionary. As a manner of storing speech part information, it may be stored as a bit train including bits of "1" and "0" representing presence and absence of speech parts as a next speech part, for example, in the descending order of the speech parts in the frequency of appearance. If 256 different speech parts are available, a storage capacity of 256 bits is required for each arc. It is to be noted that succeeding word speech part information need not be stored for all arcs.

The succeeding word speech part information is required to be stored only for an arc immediately succeeding each node as denoted by 12 in FIG. 2. This is because an arc which does not immediately succeed a node must have succeeding word speech

part information quite same as another arc immediately preceding to the arc.

By storing succeeding word speech part information only of arcs immediately succeeding nodes into the succeeding word  
5 speech part information storage section 5 in this manner, the required memory capacity can be reduced.

The succeeding word speech part information of each arc of a tree structure dictionary is required to be produced and stored only once before recognition processing is executed.  
10 This is because, when recognition processing is executed, the number of words succeeding an arc does not increase or decrease and does not change at all.

Further, also when additional registration of a word is performed to add the word to a tree structure dictionary, the  
15 succeeding word speech part information of each arc is required to be reproduced and stored only once.

In contrast, when look-ahead of a bigram is performed, each time a context is determined and a tree structure is produced, calculation of all bigram language models with  
20 respect to the context and propagation of provision of language model scores to predecessor arcs must be performed, and therefore, a very large amount of calculation is required.

The succeeding word speech part information is required to be stored by only one piece for each arc, different from  
25 a tree structure dictionary which is stored separately for each context.

This is because, if succeeding word speech part information regarding each arc on a tree structure dictionary becomes required, then it is only required to refer to the succeeding word speech part information which is stored by only one piece regarding the arc.

Where look-ahead of a bigram is performed, since it is required to store a bigram language model score for each arc of all tree structure dictionaries, a memory of a very large storage capacity is required. However, in the embodiment of the present invention, since it is only required to store a number of bit trains equal to the number of speech parts for one tree structure dictionary, the required memory capacity is reduced significantly.

Where speech parts appear in the descending order in appearance frequency of "particle", "noun", "verb" and "adjective", a bit train of such a sequence that the first bit of the speech part information of a succeeding word represents that a particle is present (1)/absent (0); the second bit represents that a noun is present (1)/absent (0), ... is stored.

It is to be noted that, although the succeeding word speech part information need not necessarily be stored in the descending order of speech parts in appearance frequency, in the following description, it is assumed that the succeeding word speech part information is stored in the descending order of speech parts in appearance frequency.

FIG. 3 illustrates an example of a bit train of succeeding

word speech part information when, where the descending order of speech parts in appearance frequency is the particle, noun, verb and adjective, succeeding words of a certain arc include a "particle", a "noun" and an "adjective", but do not include a "verb".

A speech part connection information storage section 6 stores information of all speech parts which can succeed each speech part on the right side. As a manner of storage of the speech information, for example, succeeding speech parts are stored as a bit train including bits of 1 representing that connection is allowed and 0 representing that connection is not allowed in the descending order of speech parts in appearance frequency similarly to the succeeding word speech part information storage section 5.

FIG. 4 illustrates an example of a speech part connection information bit train of a "particle" when a "particle" and a "verb" connect but a "noun" and an "adjective" do not connect to the right side of the "particle" where the descending order of the speech parts in appearance frequency is the "particle", "noun", "verb" and "adjective".

A hypothesis expansion discrimination section 2 receives a context of a tree structure dictionary on which a certain hypothesis is present from the tree structure dictionary storage section 4, receives succeeding word speech part information of an arc immediately succeeding the arc which has the hypothesis, and receives speech part connection



information of speech parts which can connect to the right sides  
of the speech part to which the context received belongs, and  
then determines from the succeeding word speech part  
information and the speech part connection information whether  
5 or not the hypothesis may be expanded to the succeeding arc.

Where the succeeding word speech part information and  
the speech part connection information are such as illustrated  
in FIGS. 3 and 4, respectively, in order to determine whether  
or not the hypothesis may be expanded to the succeeding arc,  
10 the bit train shown in FIG. 3 and the bit train shown in FIG. 4  
should be logically ANDed in a unit of a bit, and it should  
be checked whether or not a result of the operation is "0".

If the result of the operation is not "0", then this  
signifies that the hypothesis may be expanded to the succeeding  
15 arc, but if the operation result is "0", then this signifies  
that the hypothesis must not be expanded to the succeeding arc.

The reason is that, where the Nth (N is an integer equal  
to or greater than 0) bit of the speech part connection  
information is "1", only when the Nth bit of the succeeding  
20 word speech part information is "1", a word of a speech part  
whose connection is permitted is present succeeding the arc  
and the logical AND of them does not become "0".

For example, where 256 different speech parts are  
involved, if the computer has a 32-bit CPU, then  $256/32 = 8$   
25 operations are required in order to logically AND all of the  
succeeding word speech part information and the speech part

connection information.

On the other hand, where the order in which speech parts are stored is set to the descending order of the speech parts in appearance frequency, the expected value of the number of operations required to obtain a result of the logical AND which is not "0" can be suppressed smaller than 8.

In this manner, to store succeeding word speech part information and speech part connection information in the descending order of speech parts in appearance frequency contributes to high speed operation. It is to be noted that naturally it is possible to store succeeding word speech part information and speech part connection information alternatively in any other order than the descending order of speech parts in appearance frequency and perform connection checking of speech parts corresponding to each other.

It is to be noted that, if a hypothesis is expanded already to an arc immediately succeeding another arc which belongs to the hypothesis, then since it must have been determined in the past that connection is permitted, the hypothesis may be expanded to the succeeding arc unconditionally.

Based on the criterion described above, the hypothesis expansion discrimination section 2 instructs the hypothesis expansion section 3 that expansion of the hypothesis to the succeeding arc is permissible/not permissible.

The hypothesis expansion discrimination section 2 receives an instruction of expansion of the hypothesis to the

succeeding arc, acquires a structure of the arc to the succeeding arc from the tree structure dictionary storage section 4 taking scores given thereto from the language model section 7 and the acoustic model section 8 into consideration, and expands the hypothesis only to the arc, to which the hypothesis may be expanded, in accordance with a frame synchronous beam search. It is to be noted that expansion of the hypothesis to the arc itself may be performed unconditionally.

This signifies that, for example, that a hypothesis is present on a certain arc signifies that expansion of the hypothesis to the arc has been permitted in the past.

As described above, in the continuous speech recognition apparatus of the present embodiment, look-ahead of the connection possibility between a linguistic context and words in a tree structure dictionary is performed.

The present invention is effective unless exact look-ahead of a language model score more than a bigram by which the connection possibility between a context and words in a tree structure dictionary can be looked ahead, and a very large amount of calculation or memory capacity for the look-ahead of a bigram language model score is not required.

FIG. 5 illustrates operation in expansion of a hypothesis for one frame of a frame synchronous beam search.

Referring to FIG. 5, first in step S1, a loop (repetitive processing) regarding all hypotheses which are present at a

certain frame time is started.

In step S2, a context of a tree structure to which a hypothesis belongs is acquired from the tree structure dictionary storage section 4.

5 In step S3, speech part connection information of speech parts of the context is acquired from the speech part connection information storage section 6.

In step S4, an arc in the tree structure to which the hypothesis belongs is acquired from the hypothesis storage  
10 section 1. The arc acquired at this time is set as "arc 1".

In step S5, a loop regarding all succeeding arcs directly succeeding the "arc 1" is started. A directly succeeding arc selected now is set to "arc 2".

In step S6, succeeding word speech part information of  
15 the "arc 2" is acquired from the succeeding word speech part information storage section 5.

In step S7, it is determined from the speech part connection information acquired in step S3 and the succeeding word speech part information acquired in step S6 whether or  
20 not the hypothesis may be expanded from the "arc 1" to the "arc 2". If a connectable speech part included in the speech part connection information is not detected behind the "arc 2", then it is determined that the hypothesis must not be expanded to the "arc 2", and the control branches to step S9. Otherwise,  
25 it is determined that the hypothesis may be expanded to the "arc 2", and the control advances to next step S8.

In step S8 (processing by the hypothesis expansion section 3), the hypothesis is expanded to the "arc 2" in accordance with the determination in step S7.

5 In step S9, it is determined that the processing of the loop regarding all of the arcs is completed. If the processing is completed, then the control advances to step S10, but otherwise, the control returns to step S5.

10 In step S10, it is determined whether or not the processing of the loop regarding all of the hypotheses is completed. If the processing is completed, then the expansion processing of the hypotheses of the frame in the frame synchronous beam search is ended. If the processing of the loops regarding all of the hypotheses is not completed, then the control returns to step S1. It is to be noted that the  
15 processing steps described above with reference to FIG. 5 may be implemented by executing a speech recognition program (software) stored in a memory of a computer (CPU) which forms the speech recognition apparatus. In this instance, the continuous speech recognition method of the present invention  
20 can be carried out by reading out the program from a storage medium such as a floppy disc, a CD-ROM or a non-volatile memory, loading the program into the memory and executing the program by means of the CPU.

25 While a preferred embodiment of the present invention has been described using specific terms, such description is for illustrative purposes only, and it is to be understood that

changes and variations may be made without departing from the spirit or scope of the following claims.

What Is Claimed Is:

1. A continuous speech recognition apparatus,  
comprising:

5 a first storage section for storing a tree structure  
dictionary and a context preceding to the tree structure  
dictionary therein;

a second storage section for storing speech part  
information of all succeeding words of individual arcs in the  
tree structure dictionary therein;

10 a third storage section having connection rules of speech  
part information stored therein;

hypothesis expansion discrimination means for referring  
to the connection rules of speech part information stored in  
said third storage section to discriminate, with regard to each  
15 arc on which a hypothesis is present, whether or not connection  
between the context of the tree structure dictionary and a  
succeeding word of the arc is possible and determining based  
on a result of the discrimination whether or not the hypothesis  
should be expanded;

20 hypothesis expansion means for performing expansion of  
the hypothesis in response to an expansion instruction from  
said hypothesis expansion discrimination means;

whereby the hypothesis is not expanded to the arc of any  
word which does not connect to the context.

25 2. A continuous speech recognition apparatus,  
comprising:

a hypothesis storage section for storing hypotheses therein;

hypothesis expansion discrimination means for determining whether or not a hypothesis may be expanded to a succeeding arc;

a tree structure dictionary storage section for storing a tree structure dictionary and a context preceding to the tree structure dictionary therein;

a succeeding word speech part information storage section for storing information of whether or not speech parts are included in all of succeeding words present behind each of arcs in the tree structure dictionary;

a speech part connection information storage section for storing connection information between the speech parts;

means for providing a language model score to a hypothesis;

means for providing an acoustic model score to a hypothesis; and

hypothesis expansion means operable in response to an expansion instruction received from said hypothesis expansion discrimination means for acquiring a structure of an arc from said tree structure dictionary storage section and expanding a hypothesis present on the arc to a succeeding arc taking the acoustic model score and the language model score into consideration and then storing a result of the expansion into said hypothesis storage section.



3. A continuous speech recognition apparatus as claimed in claim 2, wherein said speech part connection information storage section stores connection information of the speech parts in the descending order of the speech parts in appearance  
5 frequency.

4. A continuous speech recognition apparatus as claimed in claim 2, wherein said succeeding word speech part information storage section stores the information of whether or not the speech parts are included in all of succeeding words  
10 present behind each of the arcs in the tree structure dictionary in the descending order of the speech parts in appearance frequency.

5. A continuous speech recognition apparatus as claimed in claim 2, wherein said succeeding word speech part storage  
15 section stores succeeding word speed part information regarding only those of the arcs in the tree structure dictionary each of which immediately succeeds a node.

6. A continuous speech recognition apparatus as claimed in claim 2, wherein said hypothesis expansion discrimination  
20 means receives speech part information of the context of the tree structure dictionary on which the hypothesis is being expanded from said tree structure dictionary storage section, receives speech part information of all succeeding words to an arc immediately succeeding the arc which has the hypothesis  
25 from said succeeding word speech part information storage section, refers to said speech part connection information

storage section to discriminate from the received information whether or not the context and the succeeding words can be connected to each other, and delivers, if at least one of the succeeding words is connectable to the context, an instruction  
5 to said hypothesis expansion means to expand the hypothesis to the succeeding arc, but delivers, if all of the succeeding words to the arc are not connectable to the context, an instruction to said hypothesis expansion means not to expand the hypothesis without performing expansion of the hypothesis  
10 to the succeeding arc.

7. A continuous speech recognition apparatus according to claim 2, wherein said hypothesis expansion discrimination means omits, if a hypothesis is already present on the succeeding arc to which the hypothesis is tried to be expanded,  
15 the processing of referring to the connection rules of speech parts and delivers an instruction to said hypothesis expansion means to expand the hypothesis to the succeeding arc.

8. A continuous speech recognition method, comprising the steps of:

20 storing speech part information of succeeding words to arcs in a tree structure dictionary therein;

referring to connection rules of speech part information set and stored in advance to discriminate, with regard to each arc, whether or not connection of the context of the tree  
25 structure dictionary and a succeeding word of the arc and determining based on a result of the discrimination whether

or not a hypothesis should be expanded; and

expanding the hypothesis to the arc of any word other than a word which does not connect to the context.

9. A continuous speech recognition method for a  
5 continuous speech recognition apparatus which includes a hypothesis storage section for storing hypotheses therein, a tree structure dictionary storage section for storing a tree structure dictionary and a context preceding to the tree structure dictionary therein, a succeeding word speech part  
10 information storage section for storing information of whether or not speech parts are included in all of succeeding words present behind each of arcs in the tree structure dictionary, and a speech part connection information storage section for storing connection information between the speech parts,  
15 comprising the step of:

repeating a process for all of hypotheses present at a certain frame time, the process including the steps of:

acquiring a context of a tree structure dictionary to which a hypothesis belongs from said tree structure dictionary  
20 storage section;

acquiring speech part connection information of the speech parts of the context from said speech part connection information storage section;

acquiring arcs in the tree structure dictionary to which  
25 the hypothesis belongs from said hypothesis storage section;  
and

repeating, for all succeeding arcs immediately succeeding the arcs, a process including the steps of:

acquiring, where an arc selected at present is represented as first arc and a succeeding arc immediately succeeding the first arc is represented as second arc,  
5 succeeding work speech part information of the second arc from said succeeding word speech part information storage section;

discriminating from the acquired speech part connection information and the acquired succeeding word speech part information whether or not the hypothesis may be expanded from  
10 the first arc to the second arc and determining that the hypothesis must not be expanded to the second arc if a connectable speech part included in the speech part connection information is not detected behind the second arc, but  
15 determining otherwise that the hypothesis may be expanded to the second arc;

expanding the hypothesis to the second arc; and

discriminating whether or not the loop has been completed for all of the hypotheses and ending, when the loop has been  
20 completed for all of the hypotheses, the expanding processing of the hypotheses of the frame in a frame synchronous beam search.

10. A recording medium on which a program is recorded for being executed by a computer included in a continuous speech  
25 recognition apparatus which includes a hypothesis storage section for storing hypotheses therein, a tree structure

dictionary storage section for storing a tree structure  
dictionary and a context preceding to the tree structure  
dictionary therein, a succeeding word speech part information  
storage section for storing information of whether or not  
5 speech parts are included in all of succeeding words present  
behind each of arcs in the tree structure dictionary, and a  
speech part connection information storage section for storing  
connection information between the speech parts, the program  
comprising the step of:

10 repeating a process for all of hypotheses present at a  
certain frame time, the process including the steps of:

acquiring a context of a tree structure dictionary to  
which a hypothesis belongs from said tree structure dictionary  
storage section;

15 acquiring speech part connection information of the  
speech parts of the context from said speech part connection  
information storage section;

acquiring arcs in the tree structure dictionary to which  
the hypothesis belongs from said hypothesis storage section;

20 and

repeating, for all succeeding arcs immediately  
succeeding the arcs, a process including the steps of:

acquiring, where an arc selected at present is  
represented as first arc and a succeeding arc immediately  
25 succeeding the first arc is represented as second arc,  
succeeding word speech part information of the second arc from

said succeeding word speech part information storage section;

discriminating from the acquired speech part connection information and the acquired succeeding word speech part information whether or not the hypothesis may be expanded from

5 the first arc to the second arc and determining that the hypothesis must not be expanded to the second arc if a connectable speech part included in the speech part connection information is not detected behind the second arc, but determining otherwise that the hypothesis may be expanded to

10 the second arc;

expanding the hypothesis to the second arc; and

discriminating whether or not the loop has been completed for all of the hypotheses and ending, when the loop has been completed for all of the hypotheses, the expanding processing  
15 of the hypotheses of the frame in a frame synchronous beam search.

ABSTRACT OF THE DISCLOSURE

A continuous speech recognition apparatus which does not require look-ahead of a bigram language model score which is heavy in processing. The continuous speech recognition apparatus comprises a hypothesis storage section for storing hypotheses therein, a hypothesis expansion discrimination section for determining whether or not a hypothesis may be expanded to a succeeding arc, a hypothesis expansion section for expanding a hypothesis and storing a result of the expansion into the hypothesis storage section, a tree structure dictionary storage section for storing a tree structure dictionary and a context preceding to the tree structure dictionary therein, a succeeding word speech part information storage section for storing information of whether or not speech parts are included in all of succeeding words present behind each of arcs in the tree structure dictionary, a speech part connection information storage section for storing connection information between the speech parts, a language model section for providing a language model score to a hypothesis, and an acoustic model section for providing an acoustic model score to a hypothesis. The hypothesis expansion discrimination section prevents a hypothesis from being expanded to an arc to a word which cannot connect linguistically to a context.

FIG.1

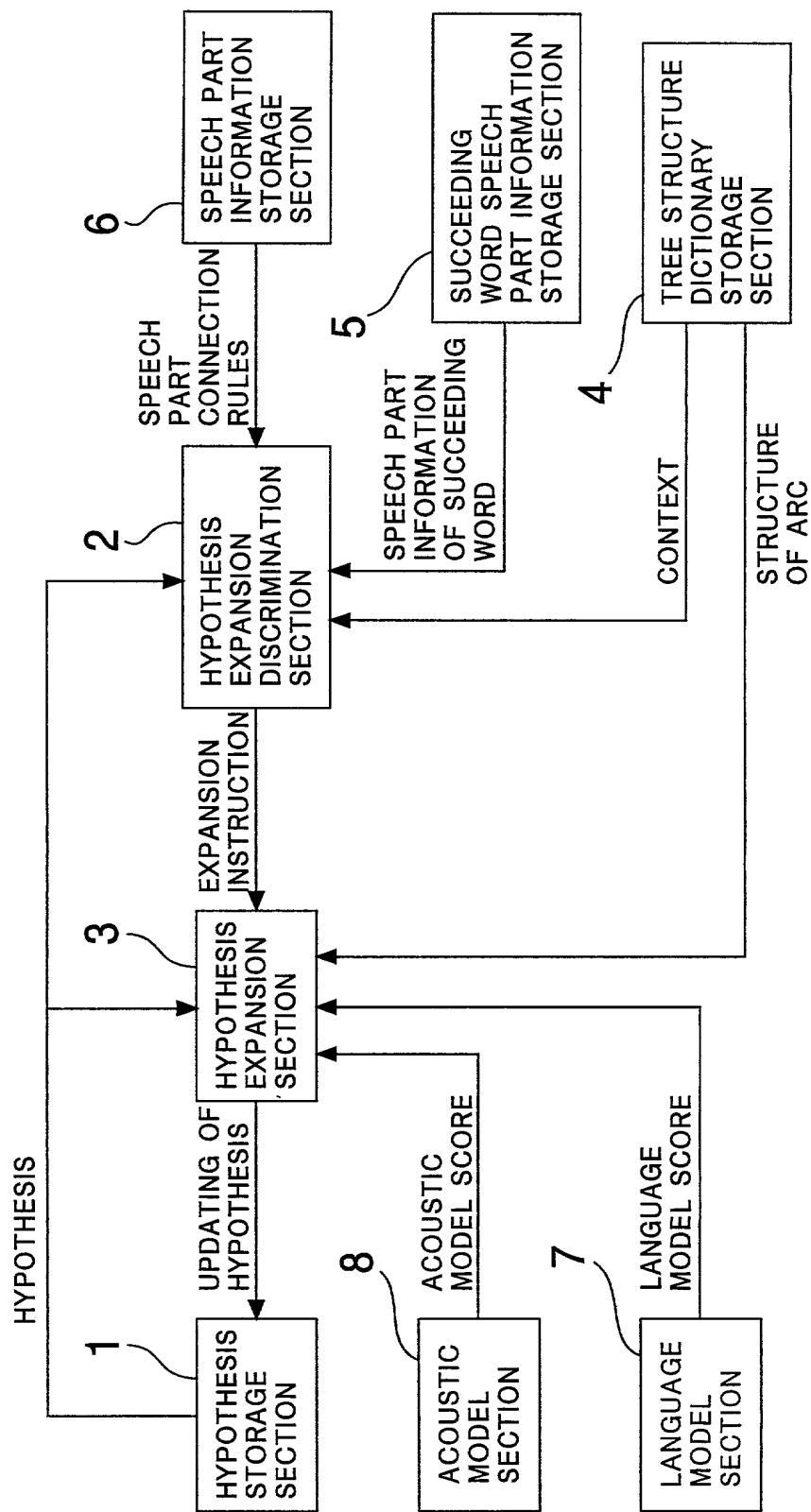




FIG.2

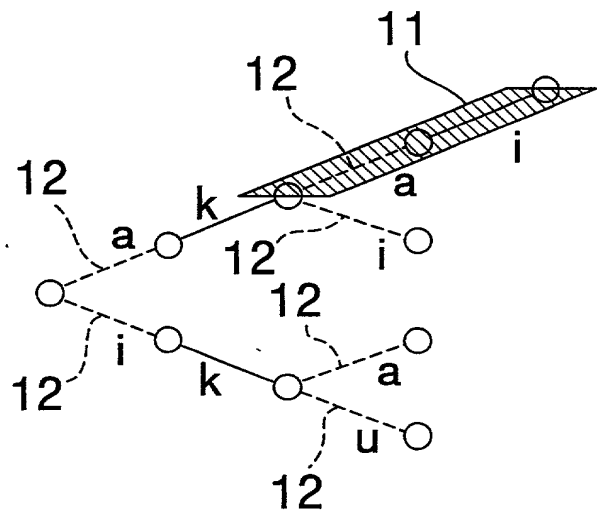


FIG.3

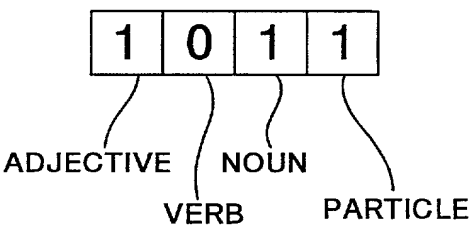


FIG.4

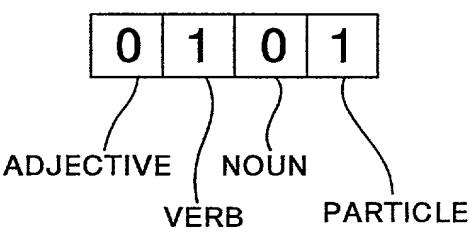


FIG.5

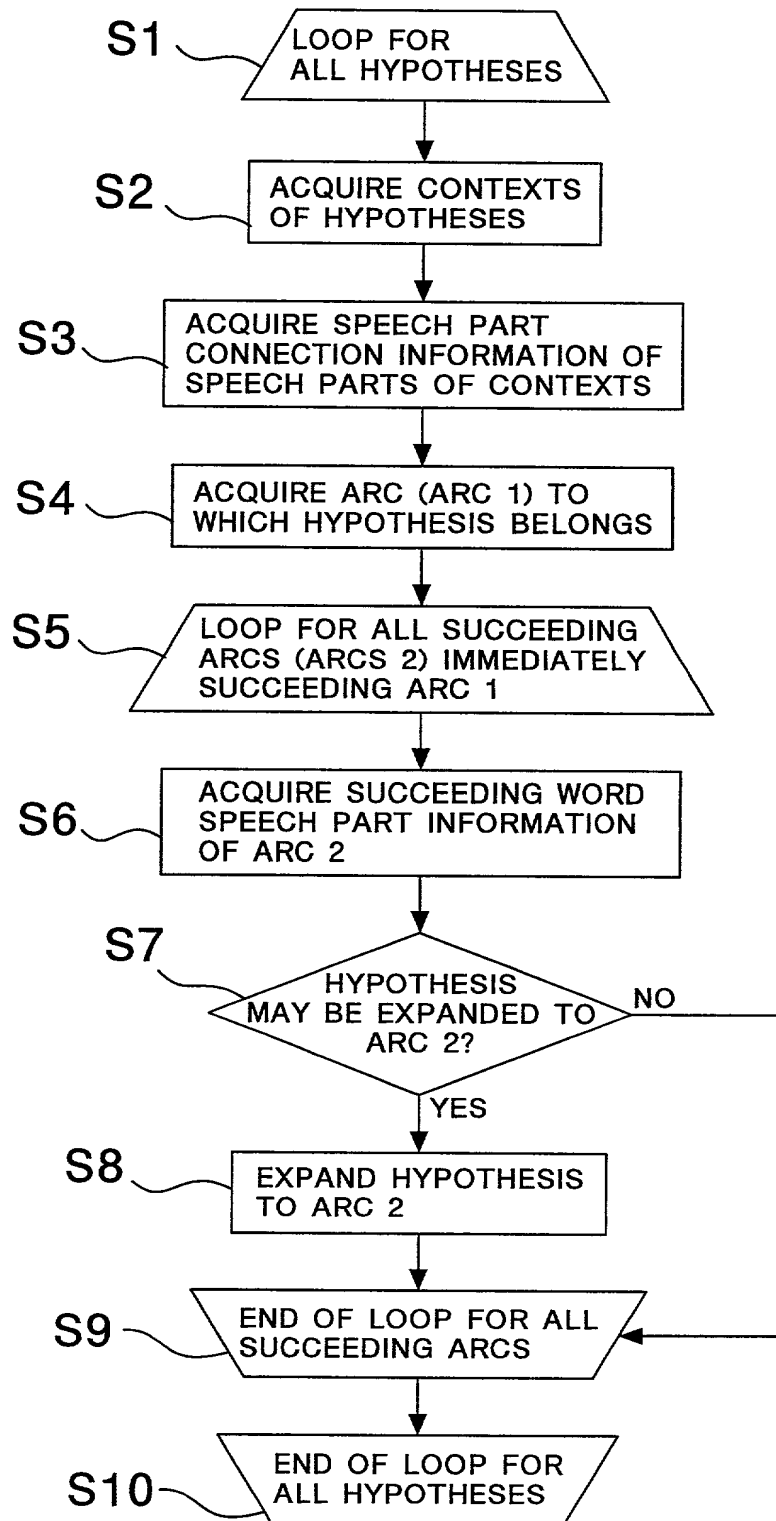
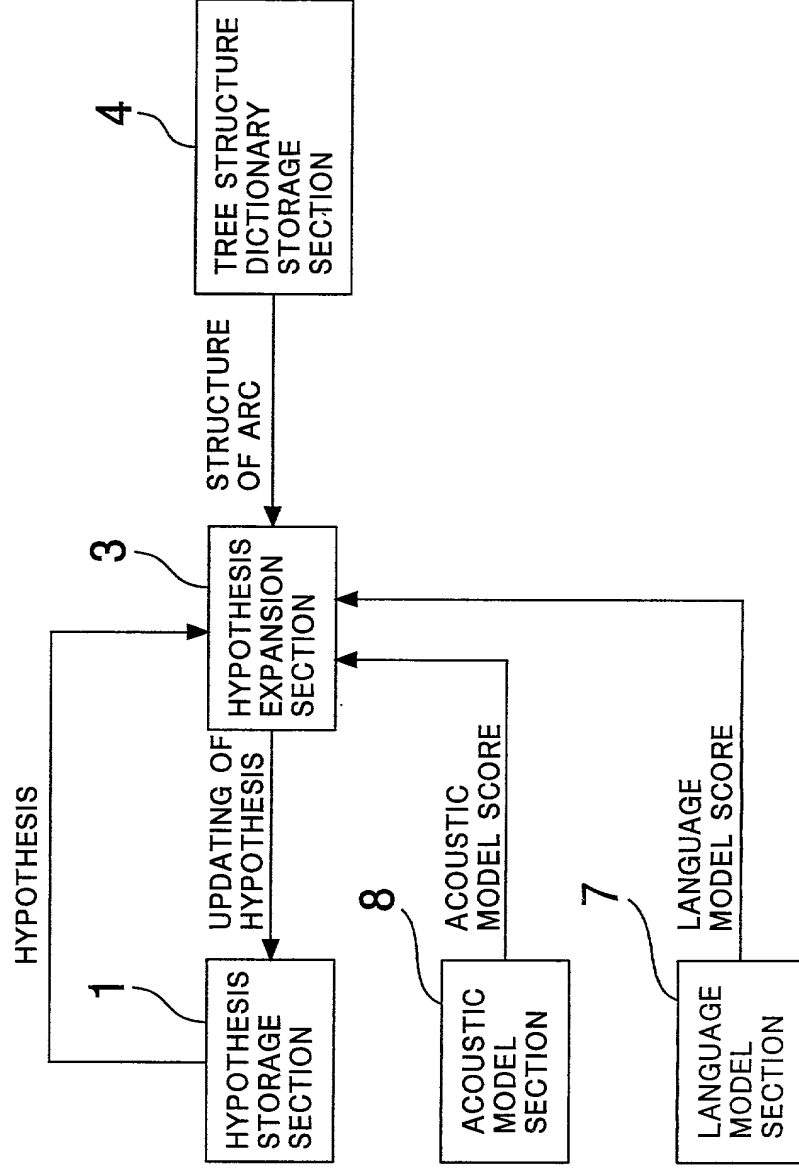


FIG.6

PRIOR ART



Docket No.

# Declaration and Power of Attorney For Patent Application

## English Language Declaration

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name,

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled

CONTINUOUS SPEECH RECOGNITION APPARATUS AND METHOD

the specification of which

(check one)

☒ is attached hereto.

☐ was filed on \_\_\_\_\_ as United States Application No. or PCT International Application Number \_\_\_\_\_ and was amended on \_\_\_\_\_ (if applicable)

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me to be material to patentability as defined in Title 37, Code of Federal Regulations, Section 1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, Section 119(a)-(d) or Section 365(b) of any foreign application(s) for patent or inventor's certificate, or Section 365(a) of any PCT International application which designated at least one country other than the United States, listed below and have also identified below, by checking the box, any foreign application for patent or inventor's certificate or PCT International application having a filing date before that of the application on which priority is claimed.

Prior Foreign Application(s)

Priority Not Claimed

<u>10-345153</u>	<u>Japan</u>	<u>4/12/1998</u>	<input type="checkbox"/>
(Number)	(Country)	(Day/Month/Year Filed)	
_____	_____	_____	<input type="checkbox"/>
(Number)	(Country)	(Day/Month/Year Filed)	
_____	_____	_____	<input type="checkbox"/>
(Number)	(Country)	(Day/Month/Year Filed)	

I hereby claim the benefit under 35 U.S.C. Section 119(e) of any United States provisional application(s) listed below:

\_\_\_\_\_  
(Application Serial No.)

\_\_\_\_\_  
(Filing Date)

\_\_\_\_\_  
(Application Serial No.)

\_\_\_\_\_  
(Filing Date)

\_\_\_\_\_  
(Application Serial No.)

\_\_\_\_\_  
(Filing Date)

I hereby claim the benefit under 35 U. S. C. Section 120 of any United States application(s), or Section 365(c) of any PCT International application designating the United States, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT International application in the manner provided by the first paragraph of 35 U.S.C. Section 112, I acknowledge the duty to disclose to the United States Patent and Trademark Office all information known to me to be material to patentability as defined in Title 37, C. F. R., Section 1.56 which became available between the filing date of the prior application and the national or PCT International filing date of this application:

\_\_\_\_\_  
(Application Serial No.)

\_\_\_\_\_  
(Filing Date)

\_\_\_\_\_  
(Status)  
(patented, pending, abandoned)

\_\_\_\_\_  
(Application Serial No.)

\_\_\_\_\_  
(Filing Date)

\_\_\_\_\_  
(Status)  
(patented, pending, abandoned)

\_\_\_\_\_  
(Application Serial No.)

\_\_\_\_\_  
(Filing Date)

\_\_\_\_\_  
(Status)  
(patented, pending, abandoned)

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith. *(list name and registration number)*

Stephen D. Murphy, Reg. No.: 22,002

Paul J. Esatto, Reg. No.: 30,749

William C. Roch, Reg. No.: 24,972


Mark J. Cohen, Reg. No.: 32,211

Frank S. DiGiglio, Reg. No.: 31,346

Donald T. Black, Reg. No.: 27,999

Send Correspondence to: Paul-J. Esatto, Jr.  
Scully, Scott, Murphy & Presser  
400 Garden City Plaza  
Garden City, New York 11530

Direct Telephone Calls to: *(name and telephone number)*  
Paul J. Esatto, Jr. (516) 742-4343

Full name of sole or first inventor	Takashi TOMOEDA	
Sole or first inventor's signature	<i>Takashi Tomoeda</i> 	Date November 29, 1999
Residence	Tokyo, Japan	
Citizenship	Japanese	
Post Office Address	c/o NEC Corporation, 7-1, Shiba 5-chome, Minato-ku, Tokyo, Japan	

Full name of second inventor, if any
Second inventor's signature
Date
Residence
Citizenship
Post Office Address